Preventive Veterinary Medicine 94 (2010) 140–153

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Combining surveillance and expert evidence of viral hemorrhagic septicemia freedom: A decision science approach

L. Gustafson [a,*], K. Klotins [b], S. Tomlinson [a], G. Karreman [b], A. Cameron [c], B. Wagner [a], M. Remmenga [a], N. Bruneau [b], A. Scott [a]

[a] USDA Animal and Plant Health Inspection Service, Veterinary Services, Centers for Epidemiology and Animal Health, 2150 Centre Ave, Bldg B, CO 80526, USA
[b] Aquatic Animal Health Division, Canadian Food Inspection Agency, 8 Colonnade Rd, Ottawa, Ontario K1A 0Y9, Canada
[c] Ausvet Animal Health Services, 4 place Puvis de Chavannes, 71480 Cuiseaux, France

ARTICLE INFO

Article history:
Received 1 December 2008
Received in revised form 24 November 2009
Accepted 25 November 2009

Keywords:
Disease freedom
Bayesian
Risk-based surveillance
Prioritization
Decision support
Evidence aggregation
Time value of information
Subjective probability

ABSTRACT

The ability to combine evidence streams to establish disease freedom or prioritize surveillance is important for the evaluation of emerging diseases, such as viral hemorrhagic septicemia virus (VHSV) IVb in freshwater systems of the United States and Canada. Waterways provide a relatively unconstrained pathway for the spread of VHSV; and structured surveillance for emerging disease in open systems has many challenges. We introduce a decision framework for estimating VHSV infection probability that draws from multiple evidence streams and addresses challenges associated with the assessment of emerging disease. Using this approach, historical and risk-based evidence, whether empirical or expert-derived, supplement surveillance data to estimate disease probability. Surveillance-based estimates of VHSV prevalence were described using beta distributions. Subjective likelihood ratios (LRs), representing contextual risk, were elicited by asking experts to estimate the predicted occurrence of risk factors among VHSV-affected vs. VHSV-unaffected watersheds. We used the odds form of Bayes' theorem to aggregate expert and surveillance evidence to predict the risk-adjusted posterior probability of VHSV-infection for given watersheds. We also used LRs representing contextual risk to quantify the time value of past surveillance data. This evidence aggregation model predicts disease probability from the combined assessment of multiple sources of information. The method also provides a flexible framework for iterative revision of disease freedom status as knowledge and data evolve.

Published by Elsevier B.V.

1. Introduction

Disease freedom is a designation that applies to populations or regions that can demonstrate, with an accepted level of confidence, a negligible likelihood of the presence of a certain disease or pathogen (OIE, 2006). Disease freedom evaluations support zonation (Zepeda et al., 2005) and trade decisions, and reduce chances of accidental pathogen transfer across management or administrative borders. However, indistinct population boundaries and the transient value of test results in systems open to new incursions complicate aquatic disease surveillance, potentially expanding sampling coverage or frequency beyond capacity of available resources. Using viral hemorrhagic septicemia virus (VHSV) as an example, we describe a decision science method adapted to combine multiple evidence streams for

* Corresponding author at: USDA Animal and Plant Health Inspection Service, Veterinary Services, Centers for Epidemiology and Animal Health, National Surveillance Unit, 2150 Centre Ave, Fort Collins, CO 80526-8117, USA. Tel.: +1 970 494 7297; fax: +1 970 494 7174.
E-mail address: lori.l.gustafson@aphis.usda.gov (L. Gustafson).

0167-5877/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.prevetmed.2009.11.021

efficient assessment of the probability and resilience of disease freedom in aquatic, open, or otherwise resource-constrained systems. Pathogen and disease are used interchangeably in this paper.

The recent emergence of a new genotype of viral hemorrhagic septicemia virus (VHSV IVb) affecting a diversity of fish species in the Great Lakes region of the United States (U.S.) and Canada (Elsayed et al., 2006; Lumsden et al., 2007) prompted emergency regulations restricting movement of susceptible fish out of the Great Lake States and provinces (http://www.aphis.usda.gov/animal_health/animal_dis_spec/aquaculture/). However, routes for water or fish passage, e.g., into Lake Champlain or the Mississippi drainage, provide natural pathways for VHSV IVb spread. Furthermore, anthropogenic movement of fish and fomites, e.g., by anglers, ballast, stocking practices or commercial trade, extend that potential to much of North America. Zonation for disease control depends on accurate delineation of diseased and disease-free regions. Consequently, U.S. and Canada bilateral surveillance was initiated to evaluate the freshwater distribution of VHSV, and to ensure that the emergent disease in freshwater fish had not spread outside of the regulated area or into aquatic animal production systems. The extensive potential taxonomic and spatial range of VHSV IVb, combined with challenges inherent to disease assessment in open systems, dictated the need to prioritize sampling for cost-effective surveillance.

However, population boundaries and sampling units for surveillance are not always clear for aquatic systems. For surveillance purposes, a watershed may be described geographically as a region whose freshwaters drain to a common destination, or by a sampling frame of its resident species- or fish-assemblages. However, natural groupings of wild freshwater fish are not always apparent to the observer, though methods for approximation of distinct population segments exist (Fransen et al., 2006). Geographic delineations, in contrast, have been defined hydrologically in both the U.S. and Canada. The freshwater drainage system in the U.S. is classified by hydrologic unit code (HUC), a U.S. Geological Survey system (http://water.usgs.gov/GIS/huc.html) used to divide the U.S. into progressively smaller geographic units defined principally by freshwater drainage patterns (Seaber et al., 1987). A similar classification system is used in Canada, though the terminology denotes primary through quaternary watersheds. In this paper, sampling units are defined geographically, with 8-digit HUCs in the U.S. comparable to tertiary watersheds in Canada.

A related challenge in disease assessment, especially of open systems, is the magnitude of sampling often required. Surveys to generate statistical evidence of disease freedom for a State might follow a 2-stage design, collecting representative samples of fish from a representative subset of the State's constituent sampling units (8-digit HUCs). Standard sample size calculations (Cameron and Baldock, 1998a,b; OIE, 2006), assuming high detection accuracy, independence of units, 95 percent confidence and a detection threshold of 10 percent prevalence, dictate sampling 27 component 8-digit HUCs (FreeCalc v.2, http://www.ausvet.com.au) for each State-level inference. From each sampled site, negative testing of 150 fish selected at random would demonstrate 95 percent probability that not more than 2 percent of fish in that sampling unit are infected with VSHV (FreeCalc v.2, http://www.ausvet.com.au) if present. Extending this approach throughout the U.S. would equate to 202,500 wild fish samples per surveillance period. This approach, however, also presumes that the collected fish adequately represent the resident fish assemblage or, if targeted sampling was used, a high-risk segment of the population. Considering the elusive nature of diseased animals in the wild, where natural loss of moribunds to predation or environmental extremes can bias routine harvests toward healthier (i.e., lower risk) fish, this assumption is not readily justified. A risk differential attributed to capture bias may, in fact, raise the need to report a detection prevalence larger than standard calculations suggest.

Further, once baseline confidence in disease freedom has been achieved, maintaining that confidence is exceedingly difficult in environments open to natural or unregulated influx of waters, migratory fish, and human and wildlife traffic. On a farm, management practices such as biosecurity protocols, facility inspections, animal import requirements, and disease monitoring programs help sustain disease freedom status. That process extends to groups of farms (e.g., by State or commodity group) if biosecurity practices, import restrictions and health regulations apply. However, these principles do not extend as readily to wild populations. Open borders raise the need for continuous or recurrent sampling, and question how often surveillance should be repeated. Consequently, population flux, indistinct boundaries and the resultant extent and frequency of sampling requirements ultimately strain both resources and the resilience of surveillance conclusions.

Aside from surveillance, however, disease status is also occasionally inferred through alternative evidence streams, such as risk assessments, historical absence of disease events, or expert opinion. The ability to credit risk-based evidence streams to optimize surveillance and support health status decisions is of recognized importance internationally (Zepeda et al., 2005; Stark et al., 2006; Martin et al., 2007a) and an ongoing focus of research and attention (Audige and Beckett, 1999; Cannon, 2002; Suess et al., 2002; Martin et al., 2007a,b; Eisler et al., 2007). With appropriate credit, these evidence streams can augment confidence derived from disease surveillance in open systems, and thereby reduce surveillance demands. However, a standardized method for assimilation of surveillance and risk-based information is essential to support judgment processes and accuracy of assessments, which may otherwise vary by decision-maker, events or time.

We describe an evidence aggregation model that facilitates risk-based prioritization of surveillance and augments previous models of disease freedom (Audige et al., 2001; Martin et al., 2007a,b). The Bayesian method, using likelihood ratios (LRs) to describe contextual evidence, has foundations in the decision sciences, e.g., to predict program or treatment success or model expert guidance for organizational change (Von Winterfeldt and

Edwards, 1986; Gustafson et al., 1992, 1993, 2003; Driver and Alemi, 1995; Bosworth et al., 1999). LRs are also used to represent the accuracy of diagnostic tests (Gallagher, 1998; Fosgate et al., 2006) or risk factors (Gustafson et al., 1998, 2005) in animal health evaluation. Epidemiologically, LRs measure the relative prevalence of a particular risk factor in diseased vs. non-diseased cohorts. LRs can be empirically derived from cross-sectional or case–control observational studies. However, elicited data, via expert opinion, is used to generate LRs when empirical data are limited in scope or quality (Von Winterfeldt and Edwards, 1986; Gustafson et al., 1992; Alemi and Gustafson, 2006). Elicited data about observable quantities or counts, carefully collected, are considered by some to be highly reliable (Gustafson et al., 1971; Alemi et al., 1986; Gingiss et al., 2006; Roberts-Gray et al., 2007) and easier to conceptualize than less observable parameters (e.g., predicted variance for a prevalence distribution) (Lele and Das, 2000; Garthwaite et al., 2005).

For surveillance planning, risk-based evidence is useful at multiple levels: both to target highest risk subjects from within a general population, and to prioritize the general population's need for surveillance by its context. Aggregating evidence for prioritization, and targeting to improve sensitivity, employ similar but distinct reasoning (Cannon, 2002; Stark et al., 2006; Martin et al., 2007a). Each relies on empirical data or expert opinion on risk. However, targeted sampling reduces the sample size required by focusing on strata (e.g., moribund vs. clinically healthy fish) within the population that are presumed most commonly affected; while evidence aggregation reduces the original need for surveillance when alternative evidence suggests a limited probability of disease in the population as a whole. Alternative evidence of disease freedom might be as definitive as a lack of susceptible species, or might simply be contributory, such as lack of known imports from VHSV infected regions. Evidence aggregation weights these contributions and asks, "Given our current state of knowledge, how much surveillance evidence do we need?" Targeting asks, "What survey design will most efficiently meet this need?" Consequently, these seemingly divergent processes are actually complementary and, applied to different stages of evaluation, together maximize surveillance efficiency.

Several methods are available to facilitate the design or evaluation of risk-based sampling for improved surveillance sensitivity; prioritization is less described (Audige and Beckett, 1999; Audige et al., 2001; Suess et al., 2002; Hadorn et al., 2002; Cannon, 2002; Branscum et al., 2005; Martin et al., 2007a,b). Surveillance sensitivity, for example, reports the confidence that disease would have been detected through surveillance if, indeed, it were present (at a given detection prevalence). Similar to negative predictive value calculations, the conditional probability that a particular watershed is truly free from disease, given the observed surveillance sensitivity results, requires also a prior estimate of disease probability (Martin et al., 2007a,b). A prior estimate ideally reflects the studied population's context, e.g., environmental or anthropogenic features presumed predictive of the probability of infection, and provides a method for juxtaposing multiple evidence streams. In practice, however, priors may be described by constructs that vary from a simple uniform distribution, to a regulatory threshold, to a distribution determined by empirical study, general intuition of the modeler, or recommendation by one or more experts (Stark et al., 2000; Audige et al., 2001; Vose, 2000; Garthwaite et al., 2005; Martin et al., 2007a).

Inaccurate or absent estimates of prior probability can seriously bias disease freedom conclusions. As an example, surveillance sensitivity statements, describing confidence about detection ability, surveillance sensitivity estimates typically exclude a prior. Derived conclusions about disease absence, then, solely reflect the system's ability to detect disease. Without consideration of context, negative results from strong surveillance in a high-risk region may suggest greater confidence in disease absence than negative results from mediocre surveillance in a region in which disease is highly implausible (e.g., a region in which susceptible species are absent). Inclusion of a prior can standardize these results. But, often a subjective assessment of a number of unspecified considerations (local knowledge of mortality events, general biosecurity conditions, the presence of susceptible species, etc.), the reasoning behind the prior probability assignment is not always clearly tracked or replicable. When surveillance data are scarce or surveillance sensitivity estimates are otherwise uncertain, the potential bias associated with an inaccurate prior becomes that much more serious.

The decision framework we describe addresses these concerns. The framework aggregates prior and current surveillance and contextual evidence, and credits expert opinion in a transparent fashion, to provide a risk-adjusted, cost-efficient, replicable and revisable estimate of disease probability. Adapted from methods designed to capture subjective probabilities and improve judgment-based decisions in medical and organizational systems (Alemi et al., 1986; Goodman, 1999; Gustafson et al., 2003), the framework uses the odds form of Bayes' theorem (Gelman et al., 2004) to support risk-based prioritization of surveillance resources and quantitative comparisons of disease probability in both surveyed and non-surveyed regions. The described model complements existing surveillance sensitivity methods such as scenario trees (Martin et al., 2007a,b) by formally placing sensitivity results in historical or environmental context. Though developed to address issues of particular relevance to aquatic populations, the method's utility extends to terrestrial diseases, especially those involving open systems.

2. Materials and methods

2.1. Sampling units

Watersheds described at the 8-digit HUC or tertiary watershed level are the epidemiologic units of the example VHSV IVb analysis. There are 21 regional or 2-digit HUCs (the entire Great Lakes basin is one regional HUC), 222 subregional or 4-digit HUCs, and 2264 cataloging unit or 8-digit HUCs in the U.S. (http://water.usgs.gov/GIS/huc.html). In Canada, the freshwater drainage system is classified according to flow volume and area through a

collaborative partnership between Water Survey of Canada (Environment Canada) and partner provincial or territorial ministries (http://www.wsc.ec.gc.ca/index_e.cfm?cname=main_e.cfm). The watershed hierarchy in Canada consists of 5 large primary or ocean watersheds, 160 secondary or river watersheds and 953 tertiary or sub-watershed watersheds. All 5 primary watersheds share portions with watersheds in the U.S.

2.2. The model structure

The evidence aggregation model is based on the odds form (Eq. (1)) of Bayes' theorem (Von Winterfeldt and Edwards, 1986; Gelman et al., 2004; Alemi and Gustafson, 2006):

$$\text{Posterior odds} = \text{prior odds} \times \text{likelihood ratios} \quad (1)$$

Prior odds represent existing knowledge about disease probability; additional evidence is represented by likelihood ratios (LRs). Prior odds and LRs derive from independent information sources, but not necessarily in temporal order (Feinberg, 1990, 2006; O'Hagan, 1998). Applied to VHSV IVb, surveillance data provide the prior odds, and expert opinion on risk factors generates a risk score (RS) representing contextual risk. The RS is the product of LRs applicable to the considered watershed. The product of the surveillance-derived odds (survey odds) and RS gives a risk-adjusted estimate of disease odds (Eq. (2)).

$$\text{VHSV posterior odds} = \text{survey odds} \times \text{RS} \quad (2)$$

Converting odds to probability generates a posterior estimate of disease probability. Following disease freedom terminology, the result is akin to the lowest detection prevalence that combined evidence streams can confidently refute. The model is amenable to parameters described by point value or probability distribution.

2.3. Surveillance evidence

Surveillance data generate a prevalence estimate or threshold detection prevalence. The prior odds of disease is derived from this prevalence estimate, where odds = prevalence/(1 − prevalence). For VHSV, prevalence is considered equivalent to the probability of disease in a randomly selected watershed. Detection prevalence is represented by a beta (a, b) distribution describing surveillance findings, where the parameters 'a' and 'b' correspond to positive and negative results, respectively, assuming independence of samples. Hydrologic units do not guarantee independence; however, they do demarcate differing degrees of functional separation and are thus an improvement over basic geographic, e.g., grid-based, units.

Representing detection prevalence in the form of a probability distribution standardizes negative results by sampling intensity. For example, negative results from a watershed with 200 sampled sites and a watershed with 1 sampled site produce equivalent mean prevalence estimates (0). However, a beta (a, b) distribution describes the range of true prevalence that might have produced that all-negative sample, thereby capturing an expected value and also the precision gained through increased sampling (Vose, 2000; Schlosser and Ebel, 2001). For the initial survey, a = positive samples + 1, b = negative samples + 1, incorporating a uniform (1, 1) prior to signify the lack of previous data. An upper bound on the distribution can be selected to more simply describe prevalence and the associated uncertainty of the estimate. For example, the 95th percentile of a beta distribution represents the prevalence level associated with 95 percent credibility. Thus, we can claim 95 percent assurance that the true prevalence of disease is less than or equal to the 95th percentile, a value dependent on the intensity (sample size) of the surveillance effort. As negative results accumulate, this 95th percentile assumes values closer to zero. The 95th percentile is a common bound for disease freedom evaluation and will be the point estimate of interest for this particular paper. However, any percentile (e.g., 95, 90 or 99 percent credibility) can be selected and identified for a given sample set using an inverse beta function found in software such as Microsoft Excel (Microsoft Office 2003, Microsoft.com).

2.4. Contextual evidence

The RS, or product of applicable LRs, represents an assemblage of contextual risk data, whether empirical or expert-derived. Watershed-level risk factors for VHSV IVb, e.g., presence of susceptible species or history of fish transfers from VHS-affected waters, predict a watershed's probability of VHSV infection. As published empirical data on risk factors and disease transmission are often limited for emerging diseases, the initial assessment of contextual risk may derive from expert opinion.

Expert opinion expressed as counts describing hypothetical data sets (e.g., predicted numbers of diseased vs. healthy animals with or without an exposure), rather than estimates for hypothetical parameters (e.g., an estimated value and variance for disease prevalence for each exposure group), is termed elicited data (Lele and Das, 2000). Asking experts to estimate the occurrence of a risk characteristic, first among a hypothetical group of cases and then among a hypothetical group of controls, provides the numerator and denominator for a subjective LR (Gustafson et al., 1992; Bosworth et al., 1999; Alemi and Gustafson, 2006). This LR represents the perceived reliability of the risk factor as a predictor of infection, essentially a ratio of the factor's perceived true positive to false positive rates. For VHSV IVb, LRs were generated by an international panel of aquatic animal health experts (VHSV Expert Panel and Working Group, this issue). Through a modified group elicitation process involving independent estimation, email discussion, revision and ultimately agreement (Gustafson et al., 1992; Bruneau et al., 1999; Bosworth et al., 1999), the panel identified nine risk factors and associated LRs perceived important to the prediction of watershed VHSV status.

Assuming conditional independence, the product of applicable risk factor LRs for any given watershed provides a summary Bayesian revision factor, here termed 'risk score' (RS), to represent contextual risk in the Bayesian model. The RS is interpreted in the same manner as LRs. A value close to one suggests that the risk factor or product of factors is neutral and has limited to no discriminative

power. A value greater or less than one suggests discriminative power as a detrimental or protective influence, respectively. Note that, as for other more common ratio measures (e.g., risk or odds ratios), extremely protective factors are bounded by zero; while extremely risky factors can conceivably approach infinity. Square-root or logistic transformations reduce the magnitude of this directional bias. Also note that an extremely strong protective factor, such as absence of susceptible species, could bring the RS close to zero and essentially negate the need for empirical surveillance data.

2.5. Historical evidence

Historic surveillance data can be combined with new surveillance data to provide a cumulative estimate of the prevalence or odds of infection. However, the relevance of past data may vary over time with risks of introduction between surveys. Intuitively, regions isolated from VHSV introduction by biosecurity or natural barriers should retain more confidence from earlier surveillance results than regions with pathways continuously open to pathogen exposure (Cannon, 2002). Key introduction pathways can be identified though import risk assessment (Hadorn et al., 2002; Stark et al., 2006), and the risk associated with each pathway estimated by the probability and recorded magnitude of introductions (Schlosser and Ebel, 2001; Hadorn et al., 2002; Alban et al., 2008). However, quantitative risk analyses and the administrative infrastructure to track data of interest (e.g., number of fish imports per year) are not always available. In lieu of more specific information, we estimated VHSV IVb introduction risk (IR) from the subset product of LRs representing introduction pathways (e.g., live fish transfers or hydrologic connectivity) relevant to the period between assessments.

The degree of temporal discounting sufficient for a given risk environment is estimated using Bayes' theorem, depreciating the strength of last period's field evidence by the likelihood of new disease introduction. Per Bayes' theorem (Eq. (1)), prior surveillance-derived odds, multiplied by an IR specific to the intervening period, produces a revised odds denoting the discounted value of prior surveillance in the current appraisal. If the region can demonstrate an absence of introduction pathways ($IR \leq 1$) since the last time of data collection, prior data are not depreciated (IR is capped at 1). If the region, however, had pathways open to new introduction (e.g., via untested fish imports) in the period between surveys ($IR > 1$), Bayesian revision inflates the odds ascribed to the previous period (revised odds = prior survey odds × IR), thus depreciating the carryover value of prior surveillance results. Odds are converted to prevalence (probability = odds/(1 + odds)); and we assign this prevalence distribution a mode of zero to represent consistently negative surveillance data. With the mode and 95th percentile, we can generate parameters for a new beta distribution (e.g., using BetaBuster software), including parameter 'b' that represents the sample size equivalent of temporally discounted data.

The observation that odds = $p/(1-p)$ can be estimated by a/b (because $p = a/n$, and $1-p = b/n$) provides an avenue for simple estimation of the IR-revised sampling equivalent of historical data.[1] Following this logic, the sampling equivalent that remains after Bayesian revision by IR is estimated by depreciated a/b_1 = original $a/b_0 \times$ IR. When $a = 1$, $b_1 = b_0/\text{IR}$, which, when rounded down to the next integer, provides the sample size equivalent of depreciated historic data. In this way, new surveillance data (n_1) combined with depreciated prior data, n_0/IR, can provide a cumulative estimated beta distribution, beta $(1, n_1 + n_0/\text{IR})$. This cumulative distribution estimates disease probability from current and historic surveillance for entry as a prior in the evidence aggregation model (Eq. (2)).

Because the IR is not structured as a rate, it is best suited for surveillance intervals of consistent or biologically justifiable length (Martin et al., 2007a). To illustrate, surveillance data collected five years prior might best be depreciated in five steps, rather than once, using a series of IRs specific to each intervening year. This helps to standardize IR-depreciated data obtained from different time periods, for example, five years vs. one year prior to the current analysis.

2.6. Evidence aggregation

The evidence aggregation model (Eq. (2)) uses Bayes' theorem to estimate posterior odds of disease through joint consideration of surveillance and contextual evidence (Fig. 1). The probability of disease freedom is represented either by a distribution or by a percentile of that distribution describing a pre-selected (e.g., 95 percent) credibility level. Survey odds of disease are estimated by surveillance data, and multiplied by the RS reflecting disease risk ascribed to contextual evidence. The product, or posterior odds, represents the risk-informed ratio of the probability of disease to the probability of disease freedom. Using this formula, it is possible to track the relative influence of the two knowledge streams on interpretation of disease freedom. Converting posterior odds to a probability (p = odds/(1 + odds)) reports results on a familiar disease freedom scale. For a pre-selected credibility level of 95 percent, a posterior probability of 2 percent implies that the combined streams of knowledge together substantiate 95 percent probability of disease freedom at a detection prevalence of 2 percent.

The same model can estimate sample size given a target posterior odds or detection prevalence. Solving Bayes' theorem (Eq. (2)) for prior odds (here, survey odds) identifies the strength of evidence required of surveillance. The risk-adjusted surveillance effort required to substantiate disease freedom is survey odds = target posterior odds/RS. The equivalent survey prevalence is odds/(1 + odds), resulting in the following statement (Eq. (3)).

$$\text{Survey prevalence} = \frac{\text{posterior odds}}{\text{RS} + \text{posterior odds}} \tag{3}$$

[1] Note that a/b produces a slightly different estimate for odds, as it more closely represents the expected value than the 95th percentile of the beta distribution. In terms of sample size, however, recommendations for samples needed to produce a desired beta distribution for prevalence (whether this curve is represented by its 95th percentile or its expected value) are equivalent.

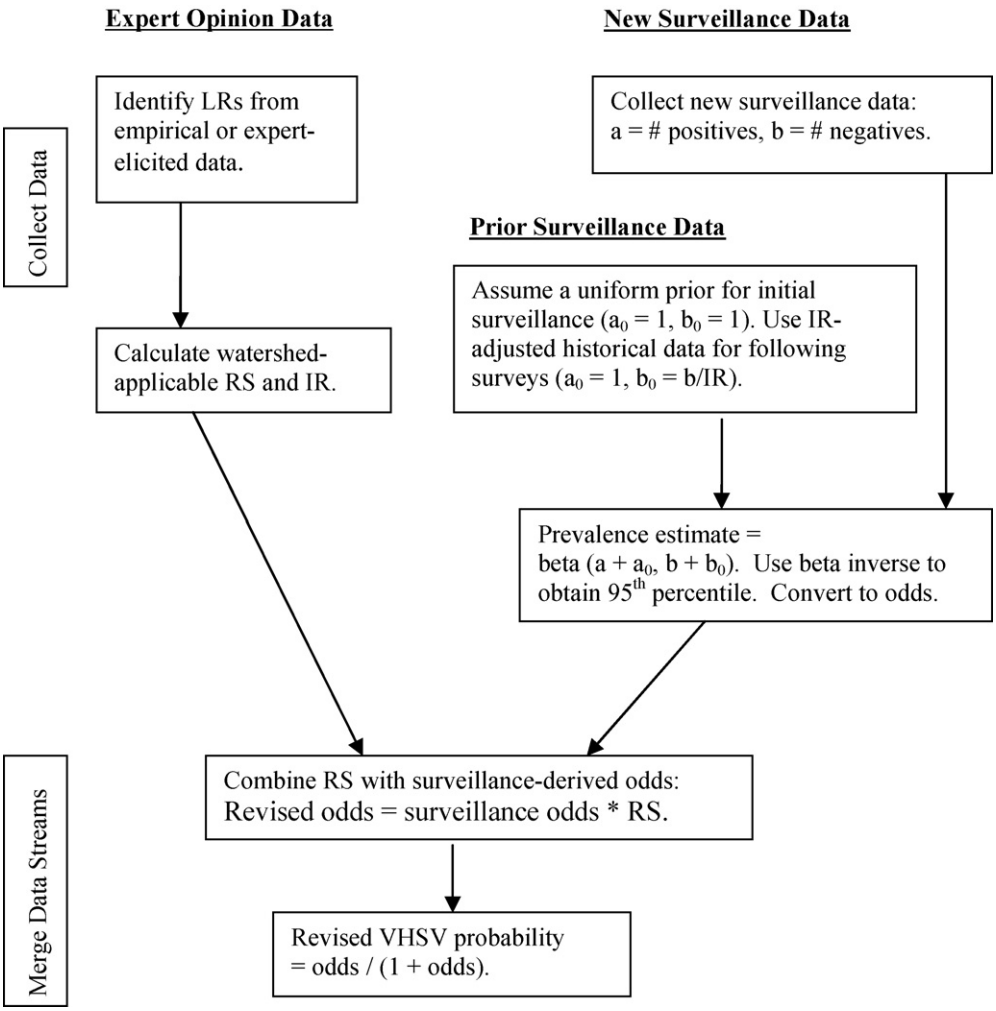

Fig. 1. Flowchart of VHSV probability estimation for initial surveillance efforts. RF denotes risk factors, LR likelihood ratios, RS risk score, and IR introduction risk. RS is the full product LRs applicable to the region of interest. IR, introduction risk, is the partial product of LRs > 1.

An RS < 1, signifying a protective context, relaxes the survey effort required to substantiate disease freedom. Assuming that we want to substantiate disease freedom for pre-determined detection prevalence, e.g., 8 percent, with 95 percent credibility, the target prevalence (8 percent) becomes the percentile assigned to that credibility level (95 percent). The necessary survey size is then determined from $b - 1$, where beta (a, b) is derived from the target percentile (here the 95th percentile is 0.08) and a mode of 0 presuming all negative findings. Again, simple approximations are justified as the odds of disease, $p/(1-p)$, can be described by beta parameters a/b (where $p/(1-p) = (a/n) \div (b/n) = a/b$). Consequently, the sampling equivalent that remains[2] after Bayesian revision by RS is readily estimated by posterior a/b_1 = survey $a/b_0 \times$ RS. When $a = 1$, which is the case for many disease freedom evaluations, this simplifies to a survey b_0 approximated by the target $b_1 \times$ RS, rounded up to the next integer. Due to resource limitations and historical precedence, RS > 1 and IR < 1 are capped at 1 to avoid inflation beyond sample sizes or temporal value considered sufficient in traditional surveillance designs.

2.7. Example model applications

To illustrate the model, we used risk scores from hypothetical watersheds to predict VHSV probability and to estimate risk-adjusted sample sizes needed to reach target confidence in disease freedom. Because limited published empirical data were available on VHSV IVb, we used risk factors and LR estimates (Table 3) generated by expert panel (VHSV Expert Panel and Working Group, this issue) to score VHSV risk (RS and IR) for each of 6 example watersheds. Our example compares two levels

[2] Note that a/b produces a slightly different estimate for odds, as it more closely represents the expected value than the 95th percentile of the beta distribution. In terms of sample size, however, recommendations for samples needed to produce a desired beta distribution for prevalence (whether this curve is represented by its 95th percentile or its expected value) are equivalent.

Table 1
95th Percentiles for beta distributions describing prevalence estimated from surveys of sample size *n*. Estimates apply to regions with no positive results and no prior data, thus assuming a uniform prior. Estimates also assume perfect sensitivity and specificity.

Survey *n*	(*a*, *b*)	95th Percentile for beta (*a*, *b*)
10	(1, 11)	0.24
20	(1, 21)	0.13
30	(1, 31)	0.09
40	(1, 41)	0.07
60	(1, 61)	0.05
80	(1, 81)	0.04
100	(1, 101)	0.03
150	(1, 151)	0.02
300	(1, 301)	0.01

of uncertainty propagation (basic and extended) using the proposed Bayesian framework.

The basic model uses point values for input and output parameters. We represented prior prevalence by the 95th percentile of a beta distribution describing surveillance findings. Here, we describe expert-derived LRs using the agreed central values (VHSV Expert Panel and Working Group, this issue). The posterior probability gives the disease prevalence that combined evidence streams can refute with a credibility of 95 percent; assuming LR point values accurately reflect expert-opinion. The full beta distribution for posterior prevalence can be generated from these results. However, since the 95th percentile is a common threshold for acceptance of confidence (in this case, credibility) in disease freedom based on results of field investigations, the extra detail provided in the full distribution adds little, if any, decision value. The basic model was constructed using an Excel spreadsheet, but the model could also be applied using a table (Table 1; showing the 95th percentile of a beta distribution for select sample sizes and results) and calculator. Note that for $n = 58$ and a beta (1, 1) prior, 1/60 (which includes the prior 'sample') is the mean of the distribution that has 0.05 as the 95th percentile.

The proposed Bayesian framework can also capture and propagate uncertainty about the LRs and associated RS. We termed the model incorporating RS uncertainty the 'extended model'. This extra layer of uncertainty was propagated through Markov chain Monte Carlo (MCMC) sampling (Audige and Beckett, 1999; Audige et al., 2001; Bruneau et al., 2001) using @Risk software (Palisade Corporation, Newfield, NY). In this version of the model, surveillance data are still represented by beta distribution, but expert-elicited LRs are described as PERT (Vose, 2000) distributions, rather than point (e.g., agreement-based) values. PERT distributions were generated from the 5th, 50th and 95th percentiles (removing the highest and lowest response) of expert-responses (square-root transformed), preceding consensus, for each LR. The output is a distribution for prevalence that incorporates both surveillance and expert-estimated LR uncertainty. This version requires simulation modeling software. Example applications of both approaches (basic and extended) are provided (Table 2).

3. Results

Evidence aggregation results for six hypothetical watersheds are shown in Tables 3–6. To estimate disease probability, RS was first calculated from the product of LRs best describing each watershed of interest. Watersheds C–F each generated a $RS < 1$, denoting some degree of protective context (Table 3). An $RS < 1$ improves confidence in disease freedom, reducing the posterior prevalence estimate beyond that achieved through surveillance alone (Table 4). As expected, uncertainty tracked in VHS probability estimates was greater in the extended vs. basic model calculations (Table 4).

The model was also used to identify risk-adjusted sample sizes for initial surveillance needs (Table 5). Given a region's computed RS, and the posterior (target) disease probability it needs to reach to demonstrate freedom, Bayes' theorem can be solved to determine the disease prevalence the region will need to refute via surveillance (Table 5). From this, the number of sample sites required can be computed using beta distributions (BetaBuster or @Risk, Palisade Corporation, Newfield, NY) or other standard methods (e.g., FreeCalc v.2, http://www.ausvet.com.au). For example, in Table 6, beta parameters for watershed D ($a = 1$ and $b = 7.5$) were estimated from a prevalence of 0.33 (the 95th percentile of the beta distribution) and a mode of zero, using BetaBuster. Based on assumptions of a uniform (1, 1) prior for the initial beta distribution, we subtract one from the sample (b) results and round up to the next whole unit to obtain sample size.

For the example watersheds, 5 percent prevalence (posterior odds = 0.05) of infection can be refuted with 95 percent credibility either by sampling and obtaining negative results from 58 sites, or by sampling just the risk-adjusted fraction of sites necessary to engender remaining confidence after consideration of contextual evidence (Table 5). Assuming 100 percent sensitivity and specificity, and an infinite population, high-risk watersheds A and B, with $RS > 1$, require full surveillance (58 sites). This is the same as not adjusting for risk. However, lower-risk regions (C–F) with $RS < 1$, receive surveillance credit for their protective context. For watershed C, with

Table 2
Function codes for model parameters. Discrete values are modeled using Excel functions. Probability distributions are modeled using @Risk functions. InvBeta provides the 95th percentile of the beta distribution. RiskBeta and PERT are @Risk functions which generate beta and PERT probability distributions, respectively. Posterior odds for the basic model are calculated as the product of prior odds and appropriate LRs. The same equation for posterior odds is solved by MCMC sampling (using @Risk) for the extended model. Parameters *a*, *b* represent positive and negative disease surveillance results, respectively; 5th and 95th provide the percentiles associated with the distribution of expert responses for LRs.

Model type	Parameter type	Prior probability (p_0)	Likelihood ratios	Posterior odds
Basic	Discrete values	InvBeta (0.95, *a*, *b*)	Consensus values	95th percentile
Extended	Distributions	RiskBeta (*a*, *b*)	RiskPERT (5th, 50th, 95th)	Full distribution

Table 3
Risk evaluation of hypothetical watersheds (WS). Explanation of risk factors and likelihood ratio (LR) derivation are provided elsewhere (VHSV Expert Panel and Working Group, this issue). Hydrologic connection, linear distance, fomite exposure and fish transfers refer to relationships with known VHSV-affected regions. Water temperatures and known susceptible species describe conditions presumed conducive to virus survival. Risk score (RS) is the product of applicable likelihood ratios for a given watershed. Introduction risk (IR) is the subset product of LRs representing open introduction pathways (excluding LRs for susceptible species and water temperature factors). RS > 1 and IR < 1 should be capped at 1 prior to use in the evidence aggregation model; an asterix (*) indicates calculated scores where this applies.

Risk factors	LR	WS A	WS B	WS C	WS D	WS E	WS F
Hydrologic connection	3.16		3.16				
	1.41						
	0.71	0.71		0.71			0.71
	0.32				0.32	0.32	
Linear distance	2.50						
	1.00	1	1			1	
	0.39			0.39	0.39		0.39
Known susceptible species	2.00	2	2	2	2	2	2
	1.22						
	0.24						
Water temperatures	1.50	1.5	1.5		1.5	1.5	
	0.47			0.47			0.47
Fomite exposure	2.24	2.24					
	1.00		1		1		
	0.39			0.39		0.39	0.39
Live fish transfer, bait	2.65	2.65		2.65			
	1.00		1			1	
	0.34				0.34		0.34
Live fish transfer, culture/stock	2.45			2.45			
	1.00	1	1		1	1	
	0.39						0.39
Frozen fish transfers	2.45				2.45		
	1.00						
	0.58	0.58	0.58	0.58		0.58	0.58
Regulatory framework	1.34	1.34					
	0.8		0.8	0.8	0.8	0.8	0.8
Risk score (RS) from PERT distributions	Median (5%, 95%)	4.57* (1.27*, 17.74*)	3.06* (0.85, 11.24*)	0.21 (0.04, 0.95)	0.24 (0.05, 1.04*)	0.11 (0.02, 0.48)	0.01 (0.00, 0.03)
Risk score (RS) from consensus LR values	Full LR product	9.83*	4.40*	0.31	0.25	0.17	0.02
Intro Risk (IR) from consensus LR values	Subset LR product	7.95	3.16	6.5	2.45	1	1

Table 4
Posterior prevalence of VHSV for 6 watersheds (WS) with varying risk scores (RS), using basic and extended models. Per Bayes' theorem, posterior odds = prior odds × RS, where odds = prevalence/(1 − prevalence). In the basic model, posterior odds are the product of discrete values for survey odds (95th percentile) and RS (accepted value). In the extended model, MCMC sampling allows uncertainty in all parameters. Beta (1, 11) represents a negative survey of 10 sites, with a uniform prior. The risk score (RS) Pert distribution is the product of applicable LR Perts (min, most likely, max), each described by the range, excluding highest and lowest values, of expert responses (Table 2 The RS Pert is capped at 1 so as not to exceed un-informed sample size requirements. MCMC sampling was set to 10,000 iterations.

WS	Survey sites	Survey prevalence		RS		Posterior prevalence			
	n	Basic	Extended	Basic	Extended	Basic	Extended		
		95th	Beta (*a*, *b*)	Agreed	Pert (5th, 50th, 95th)	95th	5th	50th	95th
A	10	0.24	(1, 11)	1	(1, 1, 1)	0.24	0.24	0.24	0.24
B	10	0.24	(1, 11)	1	(0.85, 1, 1)	0.24	0.23	0.23	0.24
C	10	0.24	(1, 11)	0.31	(0.04, 0.21, 0.95)	0.09	0.01	0.06	0.23
D	10	0.24	(1, 11)	0.25	(0.05, 0.24, 1)	0.08	0.03	0.11	0.38
E	10	0.24	(1, 11)	0.17	(0.02, 0.11, 0.48)	0.05	0.01	0.03	0.13
F	10	0.24	(1, 11)	0.02	(0.00, 0.01, 0.03)	0.01	0.00	0.00	0.01

Table 5
Example use of the VHS model for initial surveillance planning. Survey sites (sample size, *n*) required to substantiate disease freedom is calculated for 5% design prevalence and 95% credibility, for example, watersheds (WS). Surveillance required varies by risk (RS), where RS is calculated from watershed LRs (Table 2). The surveillance contribution (survey odds, SO) is calculated from prior odds = posterior odds/RS, when posterior odds (target odds, PO) describe the target prevalence detection threshold. The risk-adjusted survey prevalence (SP) is assigned to the upper credibility bound of interest (here the 95th percentile), and the mode is set to zero to represent all negative findings. Beta parameters ($a = 1$, $b = n + 1$) provide a recommended sample size *n*.

WS	Target odds (PO)	Non-adjusted sample size	RS	RS-adjusted surveillance			
				Survey odds, SO = PO/RS	Survey prevalence SP = SO/(1 + SO)	Beta (*a*, *b*)	Sample size (*n*)
A	0.05	58	1	0.05	0.05	(1, 59)	58
B	0.05	58	1	0.05	0.05	(1, 59)	58
C	0.05	58	0.31	0.17	0.15	(1, 19)	18
D	0.05	58	0.25	0.21	0.18	(1, 16)	15
E	0.05	58	0.17	0.31	0.24	(1, 11)	10
F	0.05	58	0.02	2.65	0.73	(1, 3)	2

Table 6
Incorporating risk-adjusted historical data in repeated disease freedom surveys. The following examples presume that previous efforts substantiated 95% probability of disease freedom at a 5% detection threshold. Multiplying the historical survey odds (= prior odds) by introduction risk (IR) generates the carryover value (= revised survey odds) of the original survey findings. A beta distribution for the IR-revised survey prevalence (SP) is assigned from two points: the 95th percentile set to the revised SP and a mode set to zero. Derived beta parameters ($a = 1$, $b = n$) denote the current credit ascribed to the historical data (revised sample value, *n*), rounded down to the next integer. The recommended sample size for the current planned survey is the difference between initial and revised sample size equivalents for target (Table 4) posterior detection prevalence and credibility levels.

WS	Historical data, full value		Historical data, depreciated value					Current survey
	Survey odds (SO)	Initial sample	IR	Revised survey odds	Revised survey prevalence	Beta (*a*, *b*)	Revised sample value	#Sites required
A	0.05	58	8	0.42	0.30	(1, 8.4)	8	50
B	0.05	58	3.2	0.17	0.15	(1, 18.4)	18	40
C	0.17	18	6.5	1.11	0.53	(1, 3.97)	3	15
D	0.21	15	2.5	0.53	0.35	(1, 6.95)	6	9
E	0.31	10	1	0.31	0.24	(1, 10.9)	10	0
F	2.65	2	1	2.65	0.73	(1, 2.29)	2	0

RS = 0.31, the posterior odds of 0.05 are divided by 0.31 to find that prior odds substantiated through surveillance activities would need to be 0.17 (equivalent to a prior probability of 0.15). Sample size calculations show that negative surveillance results from a sample size of 18 sites would demonstrate 95 percent probability of disease freedom at a detection prevalence of 15 percent. Thus, negative findings from a reduced ($n = 18$) set of surveillance sites, in combination with the protective (RS) contextual evidence, effectively refutes (with 95 percent credibility) disease presence at the originally targeted detection prevalence of 5 percent. Without the risk-adjustment, 58 sites would have otherwise been required.

For repeated surveillance efforts, the temporal value of historical data is estimated via the IR (Table 6). The IR (subset product of LRs that are individually > 1) represents pathways open to the risk of new disease introduction during the inter-survey period (Table 3). For VHSV, we incorporated only LRs representing true introduction pathways, i.e., excluding host and environment factors (conducive water temperatures and presence of known susceptible species), in IR calculations. Example watersheds A–D all

Table 7
Basic model equations for evidence evaluation. Evidence aggregation uses Bayes' theorem, where posterior odds (PO) = prior odds (SO) × risk score (RS). The example equations apply to regions assuming discrete LRs, pre-set credibility of 95%, and no positive findings. A beta (a, b) distribution describes prevalence. Given negative surveillance results, $a = 1$, $b = n + 1$. RS = risk score, IR = introduction risk score, LR = likelihood ratio, n = current sample value, n_0 = historic or previous sample size, n_1 = new sample size, aka = 'also known as'.

Parameter	Estimate	Caveats
RS	LR product	Cap upper bound at 1 (set RS > 1 to 1)
IR	Subset product of LRs > 1	Cap lower bound at 1 (set LR < 1 to 1)
Survey sample size, n	Sample value	Initial survey: n = actual sample size Repeat survey: $n = n_1 + n_0/\text{IR}$
Survey prevalence, SP	95th percentile of beta (1, b)[a]	Initial survey: $b = n_0 + 1$ Repeat survey: $b = n_1 + n_0/\text{IR}$
Survey odds, SO	$\text{SO} = \text{SP}/(1 - \text{SP})$	Aka prior odds in Bayes' theorem
Posterior odds, PO	$\text{SO} \times \text{RS}$	Reflects the risk-adjusted odds of disease
Posterior prevalence, PP	$\text{PO}/(1 + \text{PO})$	Reflects the risk-adjusted prevalence threshold refuted with 95% credibility

[a] See Table 8 for 95th percentiles, for example, beta parameters.

calculate IR > 1. An IR score > 1, indicating open exposure pathways, reduces the time-value of prior surveillance findings. Consequently, watersheds E and F (with an IR < 1) need only demonstrate strong biosecurity and regulatory presence to maintain a previously achieved disease-freedom status. In contrast, the other watersheds (with IR > 1) require varying levels of ongoing surveillance to retain similar confidence.

Using watershed D as an example (Table 6), we see that initial work refuted, with 95 percent credibility, disease presence at a detection prevalence of 5 percent (equivalent to posterior odds of 0.05). Both surveillance (odds = 0.21) and contextual evidence (RS = 0.25) contributed to this initial probability (Table 5), with the surveillance component (survey odds = 0.21, prevalence = 0.18) encapsulating negative results from a historical surveillance sample of size $n = 15$ (Table 6). However, the value of data from those 15 sites should depreciate over time, consequent to introduction risks present in the intervening period. For watershed D (Table 3), introduction risks are ascribed to untested imports of frozen fish from known infected regions. We impose depreciation by multiplying the initial surveillance odds (odds = 0.21) by the intervening period's IR (here 2.45), thereby estimating the current, or carryover, value (revised survey odds = 0.21 × 2.5 = 0.53, prevalence = 0.35) attributable to the previous survey. The historical IR-revised surveillance evidence (odds = 0.53, prevalence = 0.35) is now equivalent to negative results from a sample of size $n = 6$. Since 15 site equivalents are required in this generally protective risk context (RS = 0.25, Table 3) to refute disease presence at a 5% detection prevalence with 95% credibility, we now need negative findings from 9 (15 target–6 historical) new sites to maintain the region's disease freedom status.

Similarly, for watershed B (Tables 3 and 6), introduction risks are moderate (IR = 3.2) resulting in a revised posterior odds of 0.17 (from 0.05 × 3.2) or a sampling equivalent of 18 sites. In contrast, for watershed A, introduction risks are many (IR = 8), and the future value of historical information for that watershed is discounted in greater fashion, leaving a sampling equivalent of only 8 sites for carryover into the next period (Table 6). Consequently, watershed A needs to sample a greater number of sites each period to maintain confidence in disease freedom, while watershed B achieves more of its continuing confidence from prior findings.

Table 8
Basic model equations for survey planning. Evidence aggregation uses Bayes' theorem, where posterior odds (PO) = prior odds (SO) × risk score (RS). The example equations apply to regions assuming discrete LRs, pre-set credibility of 95%, and no positive findings. A beta (a, b) distribution describes prevalence. Given negative survey results, $a = 1$, $b = n + 1$. RS = risk score, IR = introduction risk score, LR = likelihood ratio, n_{dp} = standard (non-risk-adjusted) sample size for a given design prevalence, n_1 = new sample size, n_0 = historic or previous sample size. Note, design and posterior prevalence are equivalent for surveillance intended to substantiate disease freedom at a target detection threshold.

Parameter	Estimate	Caveats
RS	LR product	Cap upper bound at 1
IR	Subset product of LRs > 1	Cap lower bound at 1
Posterior prevalence, PP	Pre-set to surveillance goals	Equivalent to design prevalence
Posterior odds, PO	$\text{PO} = \text{PP}/(1 - \text{PP})$	Reflects the design prevalence
Survey odds, SO	$\text{SO} = \text{PO}/\text{RS}$	Reflects the risk-adjusted amount of evidence required of a field survey
Survey prevalence, SP	$\text{SP} = \text{SO}/(1 + \text{SO})$	Determines the risk-adjusted detection threshold required of a field survey
Beta parameter, b	Calculated from beta (1, b) describing prevalence	Beta (1, b) is solved for 95th percentile = SP and mode = 0 (using BetaBuster)
Sample size estimate, n	$n = b - 1$, for initial samples	Beta (1, b) includes a uniform prior
RS-adjusted sample size approximation	$n_1 = n_{dp} \times \text{RS}$	For initial surveys, when standard sample target (n_{dp}) is known. Round up to next integer
IR-revised sample size approximation	$n_1 = n_{dp} \times \text{RS} - n_0/\text{IR}$	For repeat surveys, when standard sample target (n_{dp}) is known. Round up to next integer

For watersheds with discrete LRs and no positive surveillance findings, the model for evidence aggregation (Table 7) and sample size approximation (Table 8) reduces to a series of basic equations that can be performed with any spreadsheet with probability distribution capabilities. Alternatively, if traditional or previous sample sizes are known, a risk-adjusted equivalent can be approximated by hand by multiplying the traditional sample size by RS, adding 1 and rounding up, or multiplying the previous sample size by 1/IR and rounding down, respectively (Tables 7 and 8). These hand-calculations lead to conservative estimates of sampling effort necessary to demonstrate or maintain disease freedom status at predetermined prevalence and credibility levels.

4. Discussion

The described evidence aggregation model, based on the odds form of Bayes' theorem, offers a quantitative framework for combining surveillance results with contextual evidence, here provided by expert opinion. The relative contributions of each evidence stream are tracked in a transparent manner; and, depicted as a series of conditionally independent LRs, the formatting of expert opinion on contextual risk is replicable and amenable to revision.

Prioritizing surveillance by contextual risk shifts sampling resources to regions or populations warranting the greatest investigative attention. Selection of sites for a risk-adjusted sampling effort can be conducted at random. However, unless the potential for disease clustering can be refuted, there is justification for purposive selection of highest risk locations. In the case of VHSV, those might include sites with high fish densities (e.g., spawning, aquaculture or enhancement) or recent mortality. If traditional or standard sample size goals are known (e.g., 60 sites to test 95 percent confidence that disease is present in less than 5 percent of the population), an estimate of the risk-adjusted sampling requirement is obtained by multiplying the uninformed sample size (e.g., 60 sites) by the region's RS, and adding 1 or rounding up to the next whole digit. If RS is 0.5, 31 negative sites would substantiate a disease prevalence of less than 5 percent. Similarly, an estimate of the sample value of historical data (e.g., 31 sites) is found by dividing the historical sample size by IR and rounding down to the next whole digit. If, for example, IR is 3, the historical data (31 sites) provide an equivalent of 10 sites to the current analysis. Assuming the RS is the same as it was for the original assessment; this implies that another 21 (31 minus 10 historic-equivalent) sites are required to maintain the population's disease freedom status in the current cycle. These simple relationships to standard sample size calculation greatly facilitate the model's field application.

Risk-based evidence can alter the interpretation of past, as well as current, surveillance data. Several approaches have been described for the time-valuation of historical surveillance data (Schlosser and Ebel, 2001; Hadorn et al., 2002; Alban et al., 2008). These methods estimate loss in data value as a function of disease introduction risks in the population over time (e.g., numbers of potentially infected animals introduced, or density-dependent contact and infectivity rates). Our described alternative, using IR to approximate disease introduction potential, is best suited for contexts in which population-based accounting measures (e.g., number of fish imported in a given period) are not available. Through Bayesian revision, the IR provides a general estimate of the potential influence of introduction pathways present. However, because IR is not structured as a rate, it is best applied to surveillance intervals of fixed and biologically justifiable (e.g., seasonal) length (Martin et al., 2007a).

The described evidence aggregation model is amenable to revision. Both RS and IR scores for the VHS application are preliminary estimates of contextual risk. As epidemiologic knowledge about a pathogen improves, LRs reflecting expert opinion can be revised or replaced with LRs based on empirical data, without impacting the underlying structure or non-disputed parameters of the model. As such, this decision support framework facilitates iterative application, with the recognition that specific factors, weights and surveillance inputs will naturally evolve. The model is designed for decision support, aggregating evidence in a transparent and comparative fashion. As such, it provides a quantitative account of the best knowledge available at a given point in time, and is thereby especially applicable to time-sensitive issues.

The model also captures uncertainty. Stochastic disease freedom models propagate uncertainty associated with input variables. The basic version of the VHS model tracks uncertainty about the surveillance-derived estimate of prevalence; the extended version also incorporates uncertainty about the LRs in the form of inter-expert variability. Alternatively, methods for approximation of confidence intervals for LRs are available (Joseph and Byorkos, 1996). However, expert response distributions reflect not only parameter variability but also methodological uncertainty, e.g., related to different heuristics influencing subjective probability assignment (Garthwaite et al., 2005). Delphi techniques narrow group uncertainty through sequential opportunities to revise individual scores after reflection on group results (Bruneau et al., 1999; Garthwaite et al., 2005). In similar format, we encouraged discussion and revision of estimates, and ultimately acceptance of the group median values (VHSV Expert Panel and Working Group, this issue).

When LR estimates are point values (e.g., acceptance-based), the basic model offers a sufficient framework for disease probability estimation. The basic version is housed in a spreadsheet with just a few functions. Furthermore, results are easily interpreted on a disease-freedom scale, and can direct future sampling to most efficiently maintain this status. This simplicity, identified as a priority in previous disease freedom classification systems (Cannon, 2002), encourages use by a broad audience. Simplicity, however, does have some drawbacks. Without the ability to also address RS uncertainty, predictions can imply a false sense of conclusion (by dismissing the importance of other sources of uncertainty).

The extended model, in contrast, allows for concurrent propagation of uncertainty in LR and RS, as well as prior probability, estimates. The full distribution of expert

estimates is captured (e.g., by a PERT distribution; Vose, 2000) via Markov Chain Monte Carlo (MCMC) sampling (Joseph and Byorkos, 1996; Audige et al., 2001) to further define uncertainty bounds on the posterior probability of disease. The ability to propagate uncertainty in all parameters makes the extended version adaptable to a broad range of data collection methods and goals. However, the complexity of MCMC sampling requires modeling capabilities outside the expected expertise of field personnel. Furthermore, a point estimate (e.g., 95th percentile) is the likely focus for many management and regulatory decisions. Consequently, the information gained through the extra model complexity, while more accurately reflecting certainty, may hold greater theoretical than practical value in many field applications.

The evidence aggregation model, as currently described, has several limitations. Conditional independence between risk factors is an assumption that allows the product of LRs to estimate an RS for use in the Bayesian model. If conditional independence of LRs is not a valid assumption, modification to include covariance terms may be necessary to improve model accuracy. Probabilistic calculation based on Bayes' theorem with strong, but possibly inaccurate, independence assumptions is sometimes referred to as 'naive Bayes'. Despite their simplistic assumptions, naïve Bayes models are known to produce remarkably accurate results for real-world classification problems (e.g., distinguishing between mutually exclusive and exhaustive states of nature) (Domingos and Pazzani, 1997). However, conditional dependencies can lead to inaccurate probability estimates, despite accurate classification.

Consequently, identifying conditional dependencies between factors is an important exercise. Accumulating data on the geographic distribution of disease and risk factors across North America may help verify the VHS model's conditional independence assumptions. Pending this external validation step, however, a preliminary check for dependence was conducted using multiple regressions of each expert's predicted VHSV risk for 42 hypothetical watersheds and associated combination of contextual risk factors (VHSV Expert Panel and Working Group, this issue). Factors lacking statistical association with expert predictions were considered redundant (potentially dependent) or minimally important to the expert's assessment of risk. These factors were either removed from the list or grouped with similar factors. However, the independence of remaining factors is ultimately assessed through external validation.

The example VHSV model, presuming 100 percent sensitivity and specificity, joined contextual risk with surveillance results directly. However, perfect accuracy is not a requirement of the model. Instead, to address greater surveillance complexity, the model structure is capable of building on system sensitivity estimates obtained from field validation trials or stochastic scenario trees (Martin et al., 2007a,b). In this instance, the design prevalence associated with the system sensitivity estimate (for example, the detection prevalence refutable with a predetermined, e.g., 95 percent, probability), whether generated via random or targeted sampling, would be entered as the surveillance-derived prevalence in the evidence aggregation model. An assessment of contextual risk, informed by expert or empirical evidence, would thereby revise this prevalence to generate a probability of disease jointly informed by both context and surveillance data.

A final, and related, caution is that the evidence aggregation model is currently designed to describe contextual risk using factors that operate at a single, and the highest, level of population inference. In our example application, the described risks operate at the level of the watershed. In practice, however, risk factors also operate at lower (e.g., animal, tank, site), and often multiple, levels. This is further complicated by the varied proportions of the population under study in each of the different risk groups, which requires proportional weighting when combining evidence streams. The current application does not account for variance in risk patterns at the sub-population level. However, we expect that future applications will find the general model structure adaptive to many of these issues. For example, scenario tree methods account for risk heterogeneity at lower population levels for system sensitivity analysis (Martin et al., 2007a). Juxtaposing contextual watershed-level risk (RS) in the current model with prevalence estimates derived through system sensitivity analysis (Martin et al., 2007a) is one way to address the risk patterns, and lack of homogeneity of disease probability, also present at lower (e.g., animal, tank, site) levels of analysis.

5. Conclusions

The evidence aggregation model is a user-friendly framework for quantitative aggregation of surveillance and contextual evidence, whether empirical or expert-derived. The model also provides a method for incorporating discounted historical surveillance data in the summary evaluation of disease status. The result is increased modeling flexibility in contexts of scarce empirical data and a potential reduction in surveillance costs, without loss in reportable confidence, in regions considered low risk. The basic version can be housed in a spreadsheet with a few functions. A beta distribution for prevalence encapsulates surveillance results; an inverse beta function gives the upper 95 percent probability limit (or other preselected percentile) of this distribution. The associated odds are multiplied by expert-derived LRs for contextual risk factors. A final step converts the revised odds of infection back to a probability. The model output depicts the lowest prevalence of disease that combined knowledge of risk and surveillance data can refute with 95 percent credibility. Alternatively, the risk-adjusted target sample size can be similarly estimated for surveillance planning purposes. Simple hand calculations using only RS or IR, and target or historic sample sizes, can mimic this process. An extended version of the model employs MCMC sampling to capture uncertainty about all model parameters, but may not offer sufficient advantage for decision-making purposes to justify the increased complexity.

The model mitigates some of the difficulties associated with disease assessment in open systems by (1) eliciting transparent and replicable opinion on contextual risk, (2)

supplementing surveillance evidence with this contextual evidence, and (3) combining the two to derive the lowest prevalence that surveillance results can credibly refute. This decision-theoretic approach frees the user from the need to estimate either a vague prior probability of infection, or a potentially arbitrary prevalence threshold defining disease absence. It provides a transparent framework for crediting information from multiple streams of evidence in disease freedom evaluation; and it is flexible to iterative review, both of inputs and results, as knowledge and data on disease accumulate. The example model application is specific to VHSV IVb. However, the method is transferable to other disease freedom investigations that elect to draw, in a replicable manner, on expert or empirical knowledge of contextual risk.

Acknowledgements

The authors thank D.H. Gustafson for foundational guidance on subjective Bayes' and decision-theoretic techniques; R.S. Bakal, P.G. Egrie, M. Farnsworth, A.E. Goodwin, P.L. Merrill, J.B. Rolland and A. Struthers for discussions that helped shape modeling strategy; and members of the VHSV Expert Panel for insights on VHSV IVb and model application.

References

Alban, L., Boes, J., Kreiner, H., Petersen, J.V., Willeberg, P., 2008. Towards a risk-based surveillance for *Trichinella* spp. in Danish pig production. Prev. Vet. Med., doi:10.1016/j.prevetmed_2008.05.008.

Alemi, F., Gustafson, D.H., Johnson, M., 1986. How to construct a subjective index. Eval. Health Prof. 9, 42–52.

Alemi, F., Gustafson, D.H., 2006. Decision Analysis for Healthcare Managers. Health Administration Press, 328 pp.

Audige, L., Beckett, S., 1999. A quantitative assessment of the validity of animal-health surveys using stochastic modeling. Prev. Vet. Med. 38, 259–276.

Audige, L., Doherr, M.G., Hauser, R., Salman, M.D., 2001. Stochastic modelling as a tool for planning animal-health surveys and interpreting screening-test results. Prev. Vet. Med. 49, 1–17.

Bosworth, K., Gingiss, P.M., Potthoff, S., Roberts-Gray, C., 1999. A Bayesian model to predict the success of the implementation of health and education innovations in school-centered programs. Eval. Prog. Plann. 22, 1–11.

Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic test sensitivity and specificity through Bayesian modeling. Prev. Vet. Med. 68, 145–163.

Bruneau, N.N., Thorburn, M.A., Stevenson, R.M.W., 1999. Use of the Delphi panel method to assess expert perception of the accuracy of screening test systems for infectious pancreatic necrosis virus and infectious hematopoietic necrosis virus. J. Aquat. Anim. Health 11, 139–147.

Bruneau, N.N., Thorburn, M.A., Wilson, J., Meek, A., 2001. Hatchery-level predictive values for infectious pancreatic necrosis virus and Aeromonas salmonicida in Ontario, Canada. Prev. Vet. Med. 48, 129–141.

Cameron, A.R., Baldock, F.C., 1998a. A new probability formula for surveys to substantiate freedom from disease. Prev. Vet. Med. 34, 1–17.

Cameron, A.R., Baldock, F.C., 1998b. Two-stage sampling in surveys to substantiate freedom from disease. Prev. Vet. Med. 34, 19–30.

Cannon, R.M., 2002. Demonstrating disease freedom—combining confidence levels. Prev. Vet. Med. 52, 227–249.

Domingos, P., Pazzani, M., 1997. On the optimality of the simple Bayesian classifier under zero-one loss. Mach. Learn. 29, 103–137.

Driver, J.F., Alemi, F., 1995. Forecasting without historical data: Bayesian probability models utilizing expert opinions. J. Med. Syst. 19, 359–374.

Eisler, M.C., Magona, J.W., Jonsson, N.N., Revie, C.W., 2007. A low cost decision support tool for the diagnosis of endemic bovine infectious diseases in the mixed crop-livestock production system of sub-Saharan Africa. Epidemiol. Infect. 135, 67–75.

Elsayed, E., Faisal, M., Thomas, M., Whelan, G., Batts, W., Winton, J., 2006. Isolation of viral haemorrhagic septicaemia virus from muskellunge, *Esox masquinongy* (Mitchill), in Lake St-Clair, Michigan, USA reveals a new sublineage of the North American genotype. J. Fish Dis. 29, 611–619.

Feinberg, S.E., 1990. Legal likelihoods and a priori assessments: what goes where? In: Geisser, S., Hodges, J., Press, S., Zellner, A. (Eds.), Bayesian and Likelihood Methods in Statistics and Econometris (Essays in Honor of George A. Barnard) North-Holland, Amsterdam, pp. 141–162.

Feinberg, S.E., 2006. Does it make sense to be an "objective Bayesian"? (comment on articles by Berger and by Goldstein). Bayesian Anal. 1, 429–432.

Fosgate, G.T., Adesiyun, A.A., Hird, D.W., Hietala, S.K., 2006. Likelihood ratio estimation without a gold standard: a case study evaluating a brucellosis c-ELISA in cattle and water buffalo of Trinidad. Prev. Vet. Med. 75, 189–205.

Fransen, B.R., Duke, S.D., McWethy, L.G., Walter, J.K., Bilby, R.E., 2006. A logistic regression model for predicting the upstream extent of fish occurrence based on geographical information systems data. North Am. J. Fish. Manage. 26, 960–975.

Gallagher, M.D., 1998. Clinical utility of likelihood ratios. Ann. Emerg. Med. 31, 391–397.

Garthwaite, P., Kadane, J., O'Hagan, A., 2005. Statistical methods for eliciting probability distributions. J. Am. Stat. Assoc. 100, 680–701.

Gelman, A., Carlin, J., Stern, H., Rubind, D., 2004. Bayesian Data Analysis, second edition. Chapman & Hall, Boca Raton, FL, 668 pp.

Gingiss, P.M., Roberts-Gray, C., Boerm, M., 2006. Bridge-It: a system for predicting implementation fidelity for school-based tobacco prevention programs. Prev. Sci. 7, 197–207.

Goodman, S.N., 1999. Toward evidence-based medical statistics. 2: The Bayes factor. Ann. Intern. Med. 130, 1005–1013.

Gustafson, D.H., Kestly, J., Greist, J., Jensen, N., 1971. Initial evaluation of a subjective Bayesian diagnostic system. Health Serv. Res. 6, 204–213.

Gustafson, D.H., Cats-bariel, W., Alemi, F., 1992. Modelling experts knowledge: the integrative group process. In: Systems to Support Health Policy Analysis: Theory, Models and Uses, Health Administration Press, Ann Arbor, MI.

Gustafson, D.H., Sainfort, F., Johnson, S., Sateia, M., 1993. The construction, reliability, validity, application, and impact of a quality index for psychiatric emergencies. Health Serv. Res. 28, 131–158.

Gustafson, D.H., Sainfort, F., Eichler, M., Adams, L., Bisognano, M., Steudel, H., 2003. Developing and testing a model to predict outcomes of organizational change. Health Serv. Res. 38, 751–776.

Gustafson, L.L., Franklin, W., Sarno, R., Hunter, R., Young, K., Johnson, W., Behl, M., 1998. Predicting early mortality of newborn guanacos by birth mass and hematological parameters: a provisional model. J. Wildl. Manage. 6, 24–35.

Gustafson, L.L., Ellis, S.K., Bartlett, C.A., 2005. Using expert opinion to identify risk factors important to infectious salmon anemia (ISA) outbreaks on salmon farms in Maine, USA and New Brunswick, Canada. Prev. Vet. Med. 70, 17–28.

Hadorn, D.C., Rufenacht, J., Hauser, R., Stark, K.D.C., 2002. Risk-based design of repeated surveys for the documentation of freedom from non-highly contagious diseases. Prev. Vet. Med. 56, 179–192.

Joseph, L., Byorkos, T.W., 1996. Inferences for likelihood ratios in the absence of a "gold standard". Med. Decis. Making 16, 412–417.

Lele, S.R., Das, A., 2000. Elicited data and incorporation of expert opinion for statistical inference in spatial studies. Math. Geol. 32, 465–487.

Lumsden, J.S., Morrison, B., Yason, C., Russell, S., Young, K., Yazdanpanah, A., Huber, P., Al-Hussinee, L., Stone, D., Way, K., 2007. Mortality event in freshwater drum Aplodinotus grunniens from Lake Ontario, Canada, associated with viral haemorrhagic septicemia virus, Type IV. Dis. Aquat. Organ. 76, 90–111.

Martin, P.A.J., Cameron, A.R., Greiner, M., 2007a. Demonstrating freedom from disease using multiple complex data sources 1: a new methodology based on scenario trees. Prev. Vet. Med. 79, 71–97.

Martin, P.A.J., Cameron, A.R., Barfod, K., Sergeant, E.S.G., Greiner, M., 2007b. Demonstrating freedom from disease using multiple complex data sources 2: case study—classical swine fever in Denmark. Prev. Vet. Med. 79, 98–115.

O'Hagan, A., 1998. Eliciting expert beliefs in substantial practical applications. Statistician 47, 21–35.

OIE World Organization for Animal Health, 2006. Manual of Diagnostic Tests for Aquatic Animals. http://www.oie.int/.

Roberts-Gray, C., Gingiss, P.M., Boerm, M., 2007. Evaluating school capacity to implement new programs. Eval. Prog. Plann. 30, 247–257.

Schlosser, W., Ebel, E., 2001. Use of a Markov-chain Monte Carlo model to evaluate the time value of historical testing information in animal populations. Prev. Vet. Med. 48, 167–175.

Seaber, P.R., Kapinos, F.P., Knapp, G.L., 1987. Hydrologic Unit Maps: U.S. Geological Survey Water-Supply Paper 2294, 63 pp., http://pubs.usgs.gov/wsp/wsp2294/.

Stark, K.D.C., Horst, H.S., Kelly, L., 2000. Combining expert opinions: a comparison of different approaches. In: Proceedings of the Ninth International Symposium on Veterinary Epidemiology and Economics, Breckenridge, USA.

Stark, K.D.C., Regula, G., Hernandez, J., Knopf, L., Fuchs, K., Morris, R.S., Davies, P., 2006. Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: review of current approaches. BMC Health Serv. Res. 6, 20–28.

Suess, E.A., Gardner, I.A., Johnson, W.O., 2002. Hierarchical Bayesian model for prevalence inferences and determination of a country's status for an animal pathogen. Prev. Vet. Med. 55, 155–171.

VHSV Expert Panel and Working Group. Viral hemorrhagic septicemia virus (VHSV IVb) risk factors and association measures derived by expert panel. Prev.Vet. Med., this issue.

Von Winterfeldt, D., Edwards, W., 1986. Decision Analysis and Behavioral Research. Cambridge University Press, New York, 624 pp.

Vose, D., 2000. Defining distributions from expert opinion. In: Risk Analysis: A Quantitative Guide, John Wiley & Sons, Chichester, England, pp. 263–290.

Zepeda, C., Salman, M., Thiermann, A., Kellar, J., Rojas, H., Willeberg, P., 2005. The role of veterinary epidemiology and veterinary services in complying with the World Trade Organization SPS agreement. Prev. Vet. Med. 67, 125–140.